

Alkuperäisen julkaisun tiedot

Keskitalo, Esa-Pekka: Aineistot arkistoidaan tulevaa käyttöä varten : Digitaalinen säilytys vaatii vain tervettä järkeä. IPRinfo 2010, nro 5. ISSN 1456-9914. Myös verkossa, http://www.iprinfo.com/lehti?action=articleDetails&a_id=827&id=51.

URN:NBN:fi-fe201102081206

Aineistot arkistoidaan tulevaa käyttöä varten - Digitaalinen säilytys vaatii vain tervettä järkeä

Esa-Pekka Keskitalo

(IPRinfo 5/2010)

Esa-Pekka Keskitalo

Pääsuunnittelija, Kansalliskirjasto

Aineistot on tuotettava ja tallennettava niin, että säilyttämisen näkökulma on mukana alusta asti.

Digitaalisia aineistoja on ollut olemassa vuosia ja vuosikymmeniä, mutta tietosisällön elinkaareissa ne ovat usein olleet vain välivaihe tai rinnakkaistuote. Kirjat on toimitettu digitaalisesti mutta painettu paperille. Verkkosanomalehdet ovat olleet otoksia painetusta lehdestä, ja asiakirjat on arkistoitu paperisina. Yhä suurempi osa informaatiosta elää kuitenkin koko elinkaarensa digitaalisena, ja digitaalisena se on myös säilytettävä. Digitaalinen säilyttäminen on uusi taiteenlaji. Se vaatii tietenkin myös investointeja, uudenlaisia tietojärjestelmiä ja ennen kaikkea uudenlaisia toimintatapoja. Digitaalinen pitkäaikaissäilytys ei kuitenkaan ole vaikeaa siinä mielessä, että olisi olemassa kallista tai salaista, harvojen hallussa olevaa huipputeknologiaa, osaamista ja erikoislaitteistoja. Sen sijaan on tarpeen suunnitella tuotannon ja tallentamisen prosessit niin, että säilymisaspekti otetaan huomioon aineiston kaikissa elinvaiheissa aina tuotantoprosessien suunnittelusta alkaen. Kysymys on lyhyesti sanoen laadun varmistamisesta.

Säilyttäminen on riskienhallintaa

Pitkäaikaissäilytys on aktiivinen tiedonhallinnan prosessi, jonka avulla varmistetaan digitaalisen aineiston käyttökelpoisuus tulevaisuudessa. Aineiston säilyvyys ei ole aineiston sisäinen ominaisuus vaan siitä huolta pitävien toimenpiteiden tulos. Näin ollen pitkäaikaissäilytyksessä ei ole kyse pelkästään varastoinnista, jossa aineiston muodostavia bittejä pyritään pitämään säilössä mahdollisimman muuttumattomana. Sellainen passiivinen säilyttäminen johtaa väistämättä aineiston käyttökelpoisuuden katoamiseen. Kyse ei myöskään voi olla jonkin tietyn tietojärjestelmän ylläpidosta, sillä niiden elinikä on vain joitakuuta vuosia. "Pitkäaikainen" tarkoittaakin tässä yhteydessä "ajanjaksoa, joka on pidempi kuin niiden ihmisten, sovellusten ja alustojen elinikä, jotka informaation alun perin loivat". Tarvitaan siis ensiksikin toimiva organisaatio, prosessit ja teknologia sekä suunnitelma siitä, miten vastuu säilytettävästä materiaalista voi ylittää erilaisia murroskohtia. Lisäksi säilyttämisessä on varauduttava myös kriiseihin. Aineiston tulee olla kohtuullisella vaivalla palautettavissa käyttökuntoon, vaikka pitkäaikaissäilytyksen prosessi välillä katkeaisikin.

Välineet vanhenevat, ihmiset hutiloivat

Digitaalisen aineiston säilymistä hankaloittaa teknologian suhteellinen vikaherkkyys. Oikein toimiakseen nykyaikaiset järjestelmät vaativat melko paljon huolenpitoa, muun muassa sähköä sekä suojaa kosteudelta, liialta ja ääriolosuhteilta. Toisaalta toimivatkin välineet vanhenevat nopeasti. Tyypillinen esimerkki on erilaiset käytöstä poistuneet muistityypit – lerppu saattaa olla priimakunnossa, mutta sen sisältämä tieto on hankala saada esille. Ja vaikka kaikki tarvittava data olisi edelleen luettavissa, sen tulkintaan tarvittavia ohjelmistoja ei kenties enää ole saatavilla, tai niitä ei voida käyttää uudistuneissa tietokoneissa ja käyttöjärjestelmissä. Pitkäaikaissäilytyksessä on siis ensinnäkin varmistettava, että data sinänsä säilyy, eli bitit pysyvät tallessa ja muuttumattomina. Toiseksi bitit täytyy säilyttää niin, että ne saadaan aina esiin; tarvittaessa ne on siirrettävä modernimpaan ympäristöön. Lisäksi bitit on saatava käyttöön. Joko käytössä tulee olla välineet lukea esimerkiksi vuonna 2011 tuotettuja Open Office -tiedostoja, tai sitten tiedostot on ajoissa muunnettu johonkin ajantasaiseen tiedostomuotoon. Pahin riski tiedon säilymiselle on kuitenkin ihminen, ennen kaikkea ihmisen huolimattomuus. Suurin osa hukkaan joutuneesta digitaalisesta tiedosta ei suinkaan ole hävinnyt kyberterrorismin tai luonnonkatastrofien seurauksena. Onpahan vain joku painanut delete-nappia ajattelemattomasti tai päästänyt säilytettävän aineiston niin sekaisin, että vaikka tieto on tallella, sen löytäminen on ylivoimaista. Ihmisen vastuulla on myös säilymistä edistävien toimenpiteiden toteuttaminen.

Aineistoon liitettävä tiedot käyttöoikeuksista

Aineistoja koskevan tiedon eli metadatan merkitys on helppo ymmärtää. Vanha valokuva voi olla sievä katsoa. Ilman tietoa siitä, ketä se esittää, sen säilyttäminen ei välttämättä ole mielekästä. Aineiston järkevä hallinta edellyttää riittävää, laadukasta ja määrämuotoista metadattaa. Säilyttämisen ja vastaisen käytön kannalta on yhtä lailla tärkeää kerätä tietoa myös aineiston teknisistä ominaisuuksista. Ne ovat tärkeitä kun suunnitellaan, millä keinoin aineiston tekninen käyttökelpoisuus taataan. Digitaalista aineistoa säilytetään käyttöä varten. Siksi on ehdottoman tärkeää kerätä metadattaa materiaalin omistushistoriasta, tekijänoikeuksista ja muista uudelleen hyödyntämisessä tärkeistä seikoista. Jos tänään syntynyttä aineistoa halutaan hyödyntää uudestaan 30 vuoden kuluttua, sen pitäisi olla mahdollista kohtuullisella vaivalla. Tällä alalla tarvitaan vielä paljon työtä. Tarvittavaa metadattaa ei vielä suinkaan aina edes oteta talteen. Esittämisen tavat kaipaavat yhteisiä käytäntöjä, jotka toimivat myös erittäin suurten datamäärien hallinnassa. Näihin asioihin onkin luvassa suosituksia.

Sisällön syntyvaiheessa pitäisi ajatella sen säilytystä

Pitkäaikaissäilytyksen onnistumiseen vaikuttavat useat seikat. Monesti ratkaisevia päätöksiä tehdään aineiston syntyvaiheessa. Synty- ja käsittelyprosesseihin liittyy usein riskitekijöitä. Aineisto saattaa olla alusta alkaen teknisesti hankalaa tai huonolaatuista. Tärkeää tietoa aineiston teknisistä ominaisuuksista, sisällöstä ja historiasta saattaa kadota. Pitkäaikaissäilyttäjän pitäisikin päästä vaikuttamaan myös aineiston syntyprosesseihin, jotta säilyttämisen edellytykset turvataan. Alussa tapahtuneita virheitä on vaikeaa, ellei mahdotonta korjata jälkikäteen. Pitkäaikaissäilytyksen näkökulma saattaa päivittäisessä työssä vaikuttaa rajoittavalta ja hidastavalta. Niin saattaa tietenkin ollakin: jos vaikkapa digitoijan on tuotettava paljon aineistoa nopeasti, on houkutus oikoa mutkissa suuri. On luonnollisesti pyrittävä löytämään realistisia toimintatapoja. On aina muistettava, kuinka kalliiksi huonoon kuntoon päässeän aineiston restaurointi tulee myös digitaalisessa maailmassa. Perustavan tärkeää on osata määritellä, mitä aineistoja säilytetään ja kuinka kauan.

Kansallinen digitaalinen kirjasto perusteilla

Kansalliskirjasto, Kansallisarkisto, Museovirasto ja suuri joukko muita digitaalista aineistoa säilyttäviä organisaatioita ovat alkaneet rakentaa yhteistä digitaalisen aineiston pitkäaikaissäilytyksen (PAS) järjestelmää. Siitä tulee osa opetus- ja kulttuuriministeriön vetovastuulla syntyvää Kansallista digitaalista kirjastoa, joka edistää kirjastojen, arkistojen ja museoiden keskeisten tietovarantojen saatavuutta ja käytettävyyttä. Vaikka hanke sinänsä on ennen kaikkea opetus- ja kulttuuriministeriön hallinnonalalla toimivien organisaatioiden välinen, sen työstä hyötyvät varmasti myös muutkin tahot. Esimerkiksi suositukset siitä, mitä metatietoja säilytettävästä aineistosta tulee kerätä ja missä muodossa, ovat varmasti kiinnostavia kaikille sisällöntuottajille. KDK-hankkeen (Kansallinen digitaalinen kirjasto) verkkosivuilla, (www.kdk2011.fi), on saatavissa runsaasti pitkäaikaissäilytyshankkeen tuottamaa materiaalia.